XXI ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
50 anos de Ciência da Informação no Brasil: diversidade, saberes e transformação social
Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

TRANSMISSÃO DE DOCUMENTOS ARQUIVÍSTICOS DIGITALIZADOS

TRANSMISSION OF DIGITIZED ARCHIVAL DOCUMENTS

Fábio Lopes de Andrade - Instituto Brasileiro de Museus (Ibram)
Cintia Aparecida Chagas - Universidade Federal de Minas Gerais (UFMG)

Modalidade: Trabalho Completo

Resumo: Este artigo é resultado de uma investigação sobre os procedimentos técnicos para a transmissão de documentos arquivísticos digitalizados a um Repositório Arquivístico Digital Confiável, pautados numa série de normativas e orientações de instituições internacionais voltadas para a preservação digital no longo prazo. O objetivo foi tornar exequível o recém editado Decreto 10.278/2020, que estabelece a técnica e os requisitos para a digitalização de documentos públicos e privados, a fim de que os documentos digitalizados produzam os mesmos efeitos legais dos documentos originais, por meio de *softwares* de código aberto, distribuição gratuita e interface amistosa ao usuário. A pesquisa é qualitativa e tem caráter exploratório. Os métodos utilizados são a pesquisa bibliográfica e documental. Os resultados alcançados indicam que é viável executar a transmissão de documentos arquivísticos digitalizados, mas que é necessário definir padrões para tarefa, como algoritmos de verificação de integridade, taxas de compressão, definição de formatos de preservação no longo prazo, definição de assinatura digital suportada, definição do subtipo de *Portable Document Format* de preservação a ser adotado utilização de *softwares* buscadores de códigos maliciosos e *softwares* para criação de pacotes de submissão de informação.

Palavras-Chave: Transmissão de documentos arquivísticos digitalizados; Preservação digital no longo prazo; Legislação Arquivística.

Abstract: *This paper investigated technical procedures for the transmission of digitized archival documents to a Trusted Digital Archival Repository, based on a series of norms and guidelines from international institutions focused on long-term digital preservation. The goal was to make feasible the Decree 10.278/2020, which establishes the technique and requirements for the digitization of public and private documents, so that the digitized documents produce the same legal effects as the original documents, through open source software, free distribution and user-friendly interface. The research was based on the exploratory methodology and bibliographic review. The results achieved indicate that it is feasible to perform the transmission of digitized archival documents, but that it is necessary to define standards for the task, such as integrity verification algorithms, compression ratio, definition of long-term preservation formats, definition of supported digital signature, definition of the Portable Document Format subtype of preservation to be adopted, and software for searching malicious code and software for creating information submission packages.*

Keywords: *Transmission of digitized archival documents; Digital preservation in the long term; Archival legislation.*

1 INTRODUÇÃO

Os documentos arquivísticos digitais diferem, significativamente, dos documentos arquivísticos em papel, afirma Rogers (2016). Eles são voláteis e sujeitos à perda, à alteração intencional ou não, à contaminação ou corrupção, mesmo quando ainda estão sob custódia de seu criador. Sua autoria, sua procedência ou cadeia de custódia podem ser difíceis ou impossíveis de determinar. Eles podem ser transmitidos, compartilhados, e copiados com facilidade. Sua acessibilidade está sujeita à obsolescência e incompatibilidade de *hardware* e *software*. Mesmo que o criador dependa de um documento arquivístico digital no curso de negócios, e mantenha sua cadeia de custódia ininterrupta, a fragilidade e vulnerabilidade exigem uma ação explícita para proteger a autenticidade do registro.

Neste contexto, a Declaração de Direitos de Liberdade Econômica, instituída pela Lei nº 13.874/2019, estabeleceu normas de proteção à livre iniciativa e ao livre exercício de atividade econômica e, em seu Art. 3º, inciso X, confere a toda pessoa natural e jurídica o direito de arquivar qualquer documento por meio digital com produção de efeitos legais e para comprovação de atos públicos (BRASIL, 2019). A regulamentação da supracitada Lei é realizada pelo Decreto 10.278/2020, que estabeleceu a técnica e os requisitos para a digitalização de documentos públicos ou privados, a fim de que os documentos digitalizados produzam os mesmos efeitos legais dos documentos originais (BRASIL, 2020).

Partindo da premissa de que os procedimentos técnicos para a transmissão de documentos arquivísticos digitalizados necessitam observar uma série de normativas e orientações arquivísticas técnicas, além de utilizar soluções de ordem tecnológica baseada em *softwares*, buscou-se, nesta pesquisa, por publicações de instituições com reconhecida experiência na preservação digital por longos períodos. Por meio de revisão bibliográfica e de testes com softwares diversos, propõe-se uma série de procedimentos com o intuito de proporcionar aos cidadãos comuns, entidades privadas e governamentais a transmissão documentos arquivísticos digitalizados com valor legal, similar ao original analógico, nas situações abrangidas pelo Decreto 10.278/2020.

2 Decreto 10.278/2020: Produtores e Consumidores

Tomando como ponto de partida o Decreto 10.278/2020, buscou-se, num primeiro momento, compreender a quem ele se destina e qual o objeto regulamentado. Assim, observando seu Art. 2º, entendeu-se que o objeto tutelado na normativa é o documento arquivístico digitalizado produzido por pessoas jurídicas de direito público interno, por pessoas jurídicas de direito privado; e por pessoas naturais (BRASIL, 2020).

Da mesma forma, julgou-se oportuno determinar quem são as pessoas jurídicas de direito interno, pessoas jurídicas de direito privado e pessoas naturais, no intuito de enquadrá-las entre as entidades do *Producer-Archive Interface Methodology Abstract Standard*[1] (PAIMAS), recomendação que define a metodologia para a estruturação das ações que necessárias entre o Produtor e o Arquivo até que os objetos de informação sejam recebidos e validados pelo Arquivo.

O PAIMAS, de acordo com Caplan, *Kehoe e Pawletko (2010), é* um padrão ISO que se baseia no *Reference model for an Open Archival Information System*[2] (OAIS) e utiliza conceitos como definidos nesse documento. Especificamente, ele elabora todas as ações e negociações que um produtor de conteúdo (Produtor) e um repositório (Arquivo) devem tomar, desde o seu contato inicial, por meio da transmissão de *Submission Information Package*[3](SIP) para um repositório, para o recebimento e validação dos SIPs pelo repositório.

Observando o Art. 41 da Lei nº 10.406/2002, que instituiu o Código Civil, temos que pessoas jurídicas de direito público interno são**:**

> I - a União;
> II - os Estados, o Distrito Federal e os Territórios;
> III - os Municípios;
> IV - as autarquias, inclusive as associações públicas;
> V - as demais entidades de caráter público, criadas por lei.
> Parágrafo único. Salvo disposição em contrário, as pessoas jurídicas de direito público, a que se tenha dado estrutura de direito privado, regem-se, no que couber, quanto ao seu funcionamento, pelas normas deste Código (BRASIL, 2002).

Da mesma forma, o Art. 44 da Lei nº 10.406/2002 define que são pessoas jurídicas de Direito Privado**:**

> I - as associações;
> II - as sociedades;

[1] Metodologia de Interface Produtor-Arquivo Padrão Abstrato.
[2] Modelo de Referência para um Sistema de Informação de Arquivo Aberto.
[3] Pacote de Submissão de Informação.

> III - as fundações;
> IV - as organizações religiosas;
> V - os partidos políticos; e
> VI - as empresas individuais de responsabilidade limitada (BRASIL, 2002).

Já pessoas naturais são definidas pelo Art. 1º da Lei nº 10.406/2002 como **"**toda pessoa que é capaz de direitos e deveres na ordem civil", de acordo com (BRASIL, 2002).

Desta feita, esta pesquisa considerou que as pessoas jurídicas de direito público interno, pessoas jurídicas de direito privado e pessoas naturais correspondem aos Produtores, partindo do entendimento do PAIMAS e do OAIS.

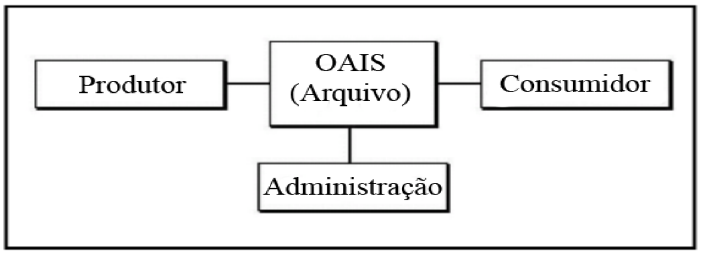
O Produtor, de acordo com Lavoie (2014, p. 9), corresponde aos "indivíduos, organizações ou sistemas que transferem informações para o OAIS para preservação no longo prazo." A negociação entre Produtores e o Arquivo OAIS especifica o conteúdo e os metadados associados que o Produtor deverá fornecer no evento de transmissão do Pacote de Submissão de Informação ao OAIS, por meio de um processo de ingestão que aceita os dados submetidos e prepara-os para a inclusão no Armazenamento de Arquivos.

Um OAIS é um Arquivo, descrito como uma "organização, que pode ser parte de uma organização maior, de pessoas e sistemas, e que aceitou a responsabilidade de preservar a informação e torná-la disponível para uma Comunidade Designada" (CCSDS, 2012)**.** Da mesma forma, considerou-se que o conceito de Arquivo, neste caso, pode ser entendido conforme se apresenta no Modelo de Referência OAIS, como sendo um Arquivo OAIS, e que se constitui dos repositórios arquivísticos digitais das instituições integrantes do Sistema Nacional de Arquivos (SINAR).

Consumidor, de acordo com CCSDS (2012), é o papel desempenhado por pessoas ou sistemas do cliente, que interagem com serviços OAIS para encontrar e adquirir informações de interesse preservadas. Partindo deste entendimento, os Consumidores podem ser considerados como qualquer cidadão, pessoa jurídica pública ou privada que deseja enviar documentos arquivísticos digitais a um ente do SINAR, entidades privadas ou pessoas.

O Ambiente circundante a um OAIS é representado na Figura 1.

FIGURA 1 - Ambiente OAIS



FONTE: CCSDS (2012, p. 2). Tradução do autor.

3 Procedimentos de responsabilidade do Produtor para a transmissão de documentos arquivísticos digitalizados

Quaisquer registros que estejam sendo transferidos para o Repositório Digital Confiável, segundo IRMT (2016), devem ser validados, colocados em quarentena e verificados para vírus no momento da ingestão. As etapas para essa parte do processo de ingestão são as seguintes:

> - Antes do *ingest*[4], os registros devem ser validados ou executados através de um *checksum*[5] que cria um resumo numérico de um registro digital (por exemplo, uma contagem do número de *bits* no registro). Isto permite que o receptor verifique se o fluxo de *bits* recebido é exatamente o mesmo que o enviado. O software de *checksum* inclui o JHOVE[6] e o Jacksum[7].
> - Outro *checksum* deve ser executado assim que os registros digitais forem transferidos para o repositório.
> - Os dois *checksums* então devem ser comparados. Se forem idênticos, o registro tem foi transmitida corretamente para o repositório digital.
> - Uma vez validados, os registros digitais precisam ser colocados em quarentena para garantir que os novos registros ingeridos não infectarão o repositório digital com nenhum vírus. Todos os registros ingeridos em um repositório digital devem ser colocados em quarentena em um servidor separado ou outra localização na rede por até 30 dias antes de serem realmente colocados no repositório. Isto é necessário para que programas de varredura de vírus atualizem seus bancos de dados, garantindo assim que todos os vírus possam ser detectados e removidos.
> - Uma vez realizada a verificação do vírus, deve ser executado um *checksum* final e comparado com o *checksum* após o *ingest*, novamente para garantir que o registro digital não tenha sido corrompido ou alterado durante qualquer um dos procedimentos de ingestão.
>
> Todas essas ações precisam ser registradas nos metadados que são ingeridos com os registros. (IRMT, 2016, p. 74, tradução dos autores).

Para Millar (2009), o processo de admissão de arquivos em um repositório digital se efetiva da seguinte maneira:

---

[4] Ingestão.
[5] Soma de verificação
[6] *JSTOR/Harvard Object Validation Environment.*
[7] *Software* que realiza somas de verificação.

> 1 Garantir que cada objeto digital a ser transferido tenha um identificador único persistente.
> 2 Verificar todos os objetos em busca de vírus e outras formas de códigos maliciosos (Esta é uma das muitas razões pelas quais é essencial garantir que o *software* antivírus esteja disponível e atualizado). Idealmente, os objetos devem ser inspecionados, colocados em quarentena por um mês e, após o final deste período, inspecionados novamente, para garantir que as viroses muito recentes sejam detectadas. Quaisquer PCs ou servidores usados para a transferência de documentos eletrônicos devem ser protegidos com programas antivírus atualizados.
> 3 Antes de transferir quaisquer registros, é necessário efetuar cópias de segurança deles, verificar sua integridade e armazená-los em uma área segura. Estes registros duplicados devem ser mantidos até que se saiba que o processo de preservação foi bem sucedido; Eles podem ser necessários como cópias-mestras caso algo dê errado com o processo de ingestão.
> 4 Uma vez que os registros tenham sido ingeridos, é necessário testar novamente os registros preservados para garantir que qualquer redução na funcionalidade, ou perda de conteúdo, estrutura ou formato, esteja dentro de limites aceitáveis. Se o processo de transferência não incluir nenhuma normalização ou outras etapas que afetem a codificação do arquivo dos componentes digitais, então um meio de validação dos registros é realizar um *checksum*. O *checksum* é executado antes e depois que os registros são transferidos a fim de confirmar que os registros não foram alterados durante a transferência. Se os registros tiverem sido corrompidos ou alterados de alguma forma, o checksum marcará o objeto digital como defeituoso.
> 5 A integridade de todos os metadados relevantes associados com os registros preservados também deve ser verificada. Em outras palavras, é importante assegurar que nenhum dos metadados foi alterado durante a transferência dos registros. Os metadados também devem ser atualizados para registrar o trabalho que tem sido feito para admitir os registros no repositório. Se a integridade dos registros não puder ser verificada, o processo de preservação terá de ser repetido em novas duplicatas dos registros da fonte. Se neste ponto o processo de ingestão ainda resulta em erros inaceitáveis, todo estratégia de preservação pode precisar ser reavaliada. (MILLAR, 2009, p. 44-45, tradução dos autores).

Baseando-se nas orientações de IRMT (2016) e Millar (2009), propõe-se a adoção de uma sequência de procedimentos abaixo descriminados, de forma a tornar exequível a construção e transmissão dos SIPs a um Repositório Arquivístico Digital Confiável (RDC-Arq) pelos entes que exerçam papel de Produtor, no contexto arquivístico e legal brasileiro.

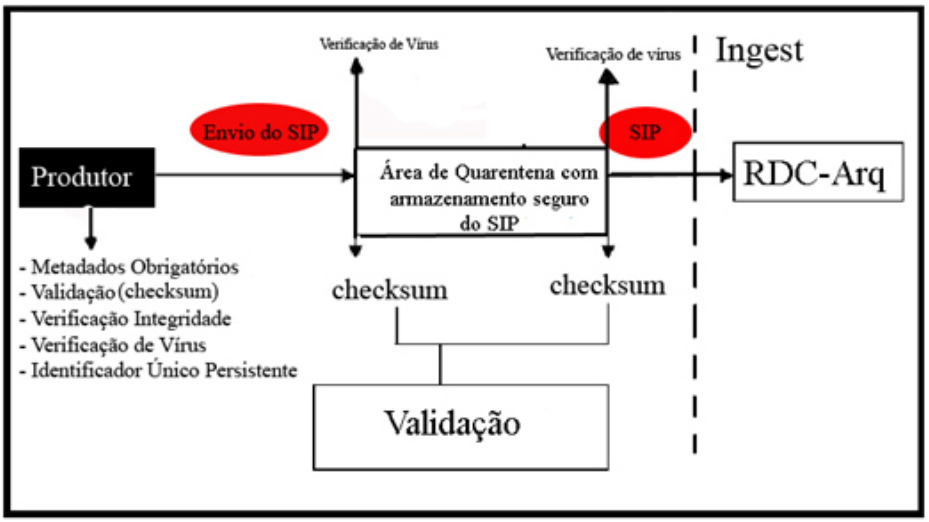
Assim, parte-se do pressuposto que o Produtor:

- Criou o representante digital (documento arquivístico digitalizado) em um ambiente de Preservação Digital Sistêmica, observando a Cadeia de Custódia Arquivística e a Cadeia de Preservação Digital do documento arquivístico;

- Observou, durante a criação do representante digital, as orientações técnicas do Anexo I do Decreto 10.278/2020, que prescreve os padrões técnicos mínimos para digitalização de documentos;

- Assinou o representante digital utilizando certificação digital no padrão da Infraestrutura de Chaves Públicas Brasileira (ICP-Brasil).

FIGURA 02 - Processo de transmissão do pacote SIP do Produtor ao RDC-Arq



FONTE: IRMT (2016, p. 75). Adaptado.

1º Procedimento: Verificação do formato do objeto digital

O Produtor deverá verificar o formato do representante digital documento arquivístico a ser transmitido e, quando necessário, convertê-lo ao formato adequado, observando o disposto no Anexo I do Decreto 10.278/2020. Os formatos de arquivo previstos no Anexo são:

- *Portable Document Format* (PDF/A); e

- *Portable Network Graphics (*PNG).

Entretanto, verificou-se, nessa pesquisa, que o Decreto 10.278/20202 não define qual subtipo de PDF/A deve ser utilizado. Para tanto recorreu-se à *Orientação Técnica nº 4 - Recomendações de uso do PDF/A para Documentos Arquivísticos* que, de acordo CONARQ (2016), apresenta recomendações gerais sobre o uso do formato PDF/A na produção e no arquivamento de documentos arquivísticos digitais.

O formato PDF/A, segundo Oettler (2013), subdivide-se em PDF/A-1, PDF/A-2 e PDF/A-3. O formato PDF/A-1 subdivide-se em PDF/A-1a e PDF/A-1b. Estes diferem entre si, principalmente, em relação aos requisitos de acessibilidade e conformidade, por exemplo.

Porém, verificou-se que a utilização de assinatura digital avançada (*PDF Advanced Electronic Signatures* - PADES), implica na utilização do formato PDF/A-2 para arquivamento de documentos, uma vez que o PDF/A-1 só permite a assinatura digital simples.

Quanto ao formato PNG, sua utilização é recomendada, segundo o Anexo I do referido Decreto, para a representação de fotografias, cartazes, plantas e mapas. Entretanto, CONARQ (2010) apresenta, além do PNG, orientações quanto à digitalização de imagens em formatos PNG, TIFF e JPEG2000. Cada um dos formatos apresenta características que, em tese, justificam a inclusão dentre os formatos que deveriam constar no Anexo I do Decreto 10.280/2020.

> - O formato *Portable Network Graphics* (PNG), apresenta como vantagem a utilização de compressão sem perdas, além, de ser um formato padronizado pela *International Standard Organization* como ISO/IEC 15948:2003. Entretanto, é mais limitado na inserção de metadados embutidos.
> - O formato mais utilizado para os representantes digitais matrizes é o formato TIFF, que apresenta elevada definição de cores sendo amplamente conhecido e utilizado para o intercâmbio de representantes digitais entre as diversas plataformas de tecnologia da informação existentes.
> - O formato de arquivo digital JPEG 2000, tem sido apreciado para a geração de matrizes quando os originais em outro formato continuam a serem preservados, mas apresenta atualmente limitações em navegação WEB, devendo ser gerada uma imagem derivada de acesso em JPEG. Pode ser configurado para fazer a compressão sem perdas. Em relação ao PNG, o JPEG 2000 permite embutir mais metadados. É um formato padronizado pela *International Standard Organization* como ISO/IEC 15444-1:2000. (CONARQ, 2010).

Para realizar a verificação dos formatos, sugere-se os *softwares Digital Record Object IDentification* (DROID) e VeraPDF para validação do Formato PDF/A. Para validação do formato PNG, sugere-se o *softwares* DROID.

A validação do formato é o processo de determinação do nível de conformidade de um objeto digital com a especificação de seu suposto formato (OPEN PRESERVATION FOUNDATION, 2015). A conformidade do formato de validação é determinada em três níveis: 1) Boa formatação, 2) Validade e 3) Consistência.

> 1) Um objeto digital é bem formatado se atender aos requisitos puramente sintáticos de seu formato;
> 2) Um objeto é válido se for bem formatado e atender aos requisitos semânticos de nível superior para a validade do formato;
> 3) Um objeto é consistente se for válido e sua informação de representação extraída internamente for consistente com a informação

> de representação fornecida externamente. (OPEN PRESERVATION FOUNDATION, 2015).

Outra ferramenta utilizada para validação do PDF/A é *JSTOR/Harvard Object Validation Environment* (JHOVE). Entretanto, surgiram dúvidas, nos últimos anos, sobre a capacidade de validação de PDF/A utilizando essa ferramenta, conforme relatam Lindlar e Tunnat (2017), Lindlar, Tunnat e Wilson (2017), que concluíram, em suas pesquisas, que o software JHOVE não deve ser utilizado para a validação de documentos em formato PDF/A, por apresentar falhas no processo de validação.

2º Procedimento: Compressão de arquivo sem perda de informação

A compressão de dados é definida, de acordo com Sharma, Naaz Mir (2018), como o processo de codificação de informações usando menos *bits* do que a representação original das informações usaria. A compressão de dados é frequentemente conhecida como *bit-rate encoding*[8] ou codificação da fonte.

Já compressão de dados sem perda envolve, segundo Berz *et al* (2015), uma transformação da representação do conjunto de dados original, de forma que é possível reproduzir exatamente o conjunto de dados originais, realizando uma transformação de descompressão. A compressão sem perda é utilizada na compressão de arquivos de texto, códigos executáveis, arquivos de processamento, arquivos de banco de dados, arquivos de tabulação, e sempre que for importante que o original e os arquivos descompactados sejam idênticos.

O Decreto 10.278/2020 orienta que, na hipótese de necessidade de comprimir o arquivo, deve ser realizada compressão sem perda, de forma que a informação obtida após a descompressão seja idêntica à informação antes de ser comprimida. Entretanto, o referido Decreto não estabelece o algoritmo de compressão a ser utilizado. Desta forma, recomenda-se, nesta pesquisa que, até a realização de experimentos mais aprofundados e a definição do melhor *software* capaz de realizar a compressão sem perda de dados, não devem ser utilizados *softwares* compressores nos SIPs a serem transmitidos a Repositórios Arquivísticos Digitais Confiáveis.

3º Procedimento: Inserção de Identificador Único Persistente

[8] Codificação de taxa de *bits*

O Identificador Único Persistente é, de acordo com IRMT (2016), um número único que facilita gestão do registro digital e o vincula com seus metadados de suporte. Registros digitais e seus metadados de suporte são frequentemente armazenados em duas áreas separadas do repositório digital. O Identificador Persistente permite que esses dois elementos permaneçam conectados, garantindo a autenticidade e a confiabilidade dos registros digitais. Exemplos de Identificadores Únicos Persistentes:

- the Uniform Resource Name (URN);
- the persistent URL (PURL);
- the Handle system;
- the digital object identifier (DOI);
- National Bibliography Numbers (NBNs);
- the Archival Resource Key (ARK);
- the Open URL (IRMT, 2016, p. 74).

Se a organização não tiver um *software* que crie um identificador persistente, um número único pode ser atribuído por um sistema de gerenciamento de registros ou de informações como uma estratégia transitória.

4º Procedimento: Inserção de Metadados obrigatórios

Verificou-se, nesta pesquisa, a necessidade da utilização de softwares para a inclusão de metadados nos SIPs. Investigamos duas ferramentas que se dispõe a isso: O BAGGER e o RODA-in. Concluiu-se que ambos os aplicativos oferecem opções como criação de checksums com diversos algoritmos, validação, edição de metadados. Entretanto, apenas o *software* RODA-in apresenta interface em língua portuguesa.

O aplicativo BAGGER foi criado para a Biblioteca do Congresso dos EUA como uma ferramenta para produzir um pacote de arquivos de dados, de acordo com a especificação BagIt. (NC DEPARTMENT OF NATURAL AND CULTURAL - NCDCR, 2019). BagIt é uma especificação, um formato hierárquico de embalagem de arquivo projetado para suportar armazenamento em disco ou em rede e transferência de arquivos digitais de conteúdo arbitrários. (ADAMS *et al*, 2018).

RODA-in é uma aplicação que pretendia produzir pacotes de submissão para o arquivo digital RODA. Os pacotes são criados *offline*, utilizando o computador do Produtor, e depois enviado para o Arquivo. Quando um novo pacote está sendo criado, o usuário deve preencher alguns campos obrigatórios a partir do esquema de metadados do *Encoded*

Archival Description[9] (EAD) como, por exemplo, o título, o nível de descrição, a referência e o produtor. Além destes campos obrigatórios, o EAD fornece campos opcionais que variam de descrições de contexto histórico para materiais associados e até descrições de conteúdo e aparência (PEREIRA, 2016).

5º Procedimento: Verificação da integridade do SIP

A verificação de integridade pode ser realizada, por meio de *checksums*. A realização de um *checksum* gera um resumo numérico de um registro digital, denominado *hash*. O registro *hash* leva em consideração, no momento de sua criação, todos os *bits* que formam o objeto digital, e esse registro será armazenado no SIP, o que permitirá que o receptor (nesse caso, o Arquivo da Instituição receptora do documento arquivístico digitalizado) verifique se o fluxo de *bits* recebido é exatamente o mesmo que o enviado. Se os registros tiverem sido corrompidos ou alterados de alguma forma durante o processo de transmissão, o *checksum* verificará que os resumos *hash* do Produtor e do Arquivo diferem e marcará o objeto digital como defeituoso.

O Decreto 10.278/2020 não estabeleceu qual o nível de complexidade de *hash* criptográfico deve ser utilizado para a verificação de integridade dos objetos arquivísticos digitalizados. Dessa forma, apresenta-se um quadro-resumo elaborado pelo *National Digital Stewardship Alliance* (NDSA) que exemplifica o nível de complexidade e esforço necessário correspondente a ser realizado pelo sistema computacional, algo que implicará em tempo demandado para a conclusão da verificação e no volume de informação processada.

QUADRO 01 - Nível de complexidade e esforço necessário correspondente realizado pelo sistema computacional

Instrumento de Fixidade	Definição	Nível de esforço e retorno sobre o investimento
Cyclic Redundancy Check (CRC)	Verificação típica de erro na rede	Baixo nível de esforço e nível moderado de detalhes. Os valores das funções CRC são variáveis mas, tipicamente, 32 ou 64 bits, que são relativamente fáceis de implementar e analisar.
Message-Digest algorithm 5 (MD5)	Função de *hash* criptográfico	Nível moderado de esforço e alto nível de detalhes. Os requisitos de CPU e processamento para calcular os valores de hash são baixos a moderados, dependendo do tamanho do arquivo. O tamanho de saída do valor de hash é o menor dos valores de hash criptográfico a 128 bits.
Secure Hash Algorithm 1 (SHA-1)	Função de *hash* criptográfico	Nível moderado de esforço, alto nível de detalhes e garantia de segurança adicional. Devido a seu maior valor de hash de saída de 160 bits, o SHA-1 requer mais tempo relativo para calcular

[9] Descrição de arquivo codificado. Tradução do autor.

		para um determinado número de ciclos de processamento CPU e tempo de processamento do que o MD5.
Secure Hash Algorithm 256 (SHA-256)	Função de *hash* criptográfico mais segura	Alto nível de esforço e muito alto nível de detalhes, e garantia de segurança adicional. Com um valor de hash de saída de 256 bits, o SHA-256 requer mais tempo relativo para calcular para um determinado número de ciclos de processamento CPU e tempo de processamento do que o SHA-1.

FONTE: National Digital Stewardship Alliance (NDSA, 2014, p. 5, tradução dos autores). Adaptado.

Os CRCs são, de acordo com NDSA(2014), úteis para gerar rapidamente informações de fixidade e são usados frequentemente no nível do conjunto de dados estruturados , no interior do registro. Entretanto, como MD5, SHA1 e SHA256 são significativamente superiores, sempre que os recursos permitem, pode ser melhor confiar em qualquer uma destas funções criptográficas de *hash* para a documentação completa de fixidade de nível de arquivo e objeto.

> Como observado acima, MD5, SHA1 e SHA256 são funções de *hash* criptográfico com diferentes tamanhos de soma de controle e com níveis crescentes de segurança. Em muitos casos, para fins de fixidez de dados, tanto MD5 quanto SHA1 são mais úteis do que SHA256 devido ao maior tempo de computação e requisitos de Unidade Central de Processamento (CPU) do computador para este último. Com o aumento dos níveis de segurança, aumenta o tempo e os recursos para calcular, portanto, dependendo da quantidade de dados em uma coleção e dos recursos disponíveis, cada um tem um lugar em diferentes fluxos de trabalho de verificação de fixidez. (NDSA, 2014, p. 6, tradução do autor).

6º Procedimento: realizar cópias de segurança dos SIPs

Registros duplicados devem ser mantidos até que se saiba que o processo de transmissão foi bem sucedido, uma vez que podem ser necessários como cópias-mestras, caso ocorram erros, durante o processo, que comprometam a integridade dos dados do pacote SIP (IRMT, 2016).

7º Procedimento: verificar o SIP quanto à existência de vírus e outras formas de códigos maliciosos

Malware é, de acordo com Monnappa (2018), um código que realiza ações maliciosas; podendo tomar a forma de um executável, *script*, código, ou qualquer outro *software*. Os invasores usam *malware* para roubar informações sensíveis, espionar o sistema infectado, ou assumir o controle do sistema. Normalmente, ele entra em seu sistema sem

consentimento do usuário e pode ser entregue, por meio de vários canais de comunicação, tais como e-mail, *web*, ou *drives* USB.

8º Procedimento: verificar as formas para a transmissão do SIP

É importante que o Produtor verifique a capacidade que a operadora de internet contratada disponibiliza para velocidade de conexão, bem como sua estabilidade no momento de transmissão, principalmente quando o pacote de informação alcançar um volume considerável (Sugestão dos autores).

3 CONSIDERAÇÕES FINAIS

Diante dos esforços investigativos empreendidos acima, pode-se concluir que o processo de transmissão de um SIP para um RDC-Arq é bastante complexo e implica numa série de ações executadas por diferentes *softwares*, mas com um objetivo único, ao final: garantir a correspondência entre os *bits* fornecidos pelo Produtor e o fluxo de *bits* recebido pelo RDC-Arq, apoiados por evidências baseadas em critérios matemáticos, através de algoritmos utilizados como artifício para verificação de integridade. Entretanto, é imperativo que a legislação arquivística brasileira que trata do assunto debata com a comunidade de pesquisa científica arquivística questões como:

- Subtipos do formato de PDF/A a serem adotados, levando em conta questões como *PDF Advanced Electronic Signatures*[10] (PaDES) e Optical Character Recognition [11](OCR);

- Acréscimo dos formatos TIFF e JPEG 2000 dentre os formatos que representem imagens estáticas, no Anexo I do Decreto 10.278/2020;

- Definição do algoritmo de compressão sem perdas a ser utilizado;

- Definição do nível de esforço para a criação do *hash* criptográfico;

- Identificadores Únicos Persistentes para documentos arquivísticos digitais.

- Detecção de códigos maliciosos embutidos em PDF/A;

REFERÊNCIAS

ADAMS, C. *et al.* **The BagIt file packaging format (V1.0)**. Califórnia: IETF Trust, 2018. Disponível em: https://tools.ietf.org/html/rfc8493^. Acesso em: 18 jan. 2021.

BERZ, Dominic *et al.* Comparison of lossless data compression methods. **Technical Reports in Computing Science**, Kempten, n. CS-07, p. 1-12, 2015.

[10] Assinatura digital avançada. Tradução do autor.
[11] Reconhecimento óptico de caracter. Tradução do autor.

BRASIL. Decreto 10.278 de 18 de março de 2020. Regulamenta o disposto no inciso X do caput do art. 3º da Lei nº 13.874, de 20 de setembro de 2019, e no art. 2º-A da Lei nº 12.682, de 9 de julho de 2012, para estabelecer a técnica e os requisitos para a digitalização de documentos públicos ou privados, a fim de que os documentos digitalizados produzam os mesmos efeitos legais dos documentos originais. **Diário Oficial da União**, DF, mar. 2020. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/decreto/D10278.htm. Acesso em: 15 jan. 2021.

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002**. Institui o código civil. Brasília: Presidência da República, 2002. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/2002/L10406compilada.htm. Acesso **em:** 15 jan. 2021.

BRASIL. Lei nº 13.874, de 20 de setembro de 2019. Institui a Declaração de Direitos de Liberdade Econômica; estabelece garantias de livre mercado; altera as Leis nºs 10.406, de 10 de janeiro de 2002 (Código Civil), 6.404, de 15 de dezembro de 1976, 11.598, de 3 de dezembro de 2007, 12.682, de 9 de julho de 2012, 6.015, de 31 de dezembro de 1973, 10.522, de 19 de julho de 2002, 8.934, de 18 de novembro 1994, o Decreto-Lei nº 9.760, de 5 de setembro de 1946 e a Consolidação das Leis do Trabalho, aprovada pelo Decreto-Lei nº 5.452, de 1º de maio de 1943; revoga a Lei Delegada nº 4, de 26 de setembro de 1962, a Lei nº 11.887, de 24 de dezembro de 2008, e dispositivos do Decreto-Lei nº 73, de 21 de novembro de 1966; e dá outras providências. **Diário Oficial da União**, DF, set. 2019. **Disponível em:** http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/L13874.htm**. Acesso em:** 15 jan. **2021.**

CAPLAN, Priscilla; KEHOE, William; PAWLETKO, Joseph. Towards Interoperable Preservation Repositories (TIPR). **International Journal of Digital Curation**, Bath, v. 5, n. 1, p. 34-45, mar. 2010.

CONSULTIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). **Reference model for an Open Archival Information System (OAIS).** Washington: CCSDS Secretaria, 2012. 135p.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Câmara Técnica de Documentos Eletrônicos. **Orientação Técnica nº 4, outubro de 2016.** Rio de Janeiro: CONARQ, 2016. 13p.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). Resolução nº 31, de 28 de abril de 2010. Dispõe sobre a adoção das Recomendações para Digitalização de Documentos Arquivísticos Permanentes. **Diário Oficial da União**, DF, 3 maio 2010. Seção 1, nº 82.

CONSULTIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS**)**. **Producer-archive interface methodology abstract standard**. Washington: CCSDS Secretaria, 2004b. 72p.

DIGITAL PRESERVATION (iPRES2017), 14., 2017, Kyoto. **Proceedings…** Kyoto: iPRES, 2017. p. 1-11.

DIGITAL PRESERVATION COALITION (DPC). **Fixity and checksums**. Glasgow: DPC, 2020. Disponível em: https://www.dpconline.org/handbook/technical-solutions-and-tools/fixity-and-checksums. Acesso em: 15 jan. 2021.

INTERNATIONAL RECORDS MANAGEMENT TRUST (IRMT). **Digital preservation in lower resource environments:** a core curriculum. Paris: ICA, 2016. 94p.

LAVOIE, Brian F. **The Open Archival Information System (OAIS) Reference Model:** introductory guide. 2. ed. Glasgow: DPC, 2014. 33p.

LINDLAR, Michelle; TUNNAT, Yvonne. How valid is your validation? A closer look behind the curtain of JHOVE. **International Journal of Digital Curation**, Bath, v. 12, n. 2, p. 286-298, 2017.

LINDLAR, Michelle; TUNNAT, Yvonne; WILSON, Carl. **A PDF Test-Set for Well Formedness Validation in JHOVE** - The Good, the Bad and the Ugly. *In*: INTERNATIONAL CONFERENCE ON

MILLAR, Laura (ed.). **Module 4:** preserving electronic records. London: IRMT, 2009. 57p.

MONNAPPA, K. A. **Learning malware analysis:** explore the concepts, tools, and techniques to analyze and investigate Windows malware. Birmingham: Packt Publishing, 2018. 512p.

NATIONAL DIGITAL STEWARDSHIP ALLIANCE (NDSA). **Checking your digital content:** how, what and when to check fixity?. Arlington: NDSA, 2014. 7p.

NC DEPARTMENT OF NATURAL AND CULTURAL RESOURCES. **Bagger GUI user guide:** how to create and validate Bags with Bagger. Estados Unidos: NC, 2019. 25p.

OETTLER, Alexandra**. PDF/A *in a Nutshell 2.0****: PDF for long-term archiving. The ISO Standard – from PDF/A-1 to PDF/A-3.* Association for Digital Document Standards. 2013 Berlin.

OPEN PRESERVATION FOUNDATION. **JHOVE. Software.** Leeds: Open Preservation Foundation, 2015. Disponível em: http://jhove.openpreservation.org/. Acesso em: 15 jan. 2021.

PEREIRA, Andre. **RODA-in** - a generic tool for the mass creation of submission information packages. Masters dissertation. Masters in Informatics Engineering. Escola de Engenharia. Departamento de Informática. Universidade do Minho. 2016. 74p.

ROGERS, Corinne. A literature review of authenticity of records in digital systems from 'machine-readable' to records in the cloud. **Acervo**, Rio de Janeiro, v. 29, n. 2, p. 16-44, jul./dez. 2016.

SHARMA, Vipul; NAAZ MIR, Roohie. Digital preservation and data compression. **International Journal of Computer Science and Technology**, London, v. 9, n. 2, p. 32-43, Apr./June 2018.

THE NATIONAL ARCHIVES. **The technical registry PRONOM**. [*S.l.*: *s.n.*], 2021. Disponível em: www.nationalarchives.gov.uk/help/PRONOM/faq.htm#faq1. Acesso em: 15 jan. 2021.

VeraPDF. **Desktop GUI quick start guide**. 2020. Disponível em: https://docs.verapdf.org/gui/. Acesso em: 15 jan. 2021.